AD-A258 166 **ON PAGE**

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
13 Oct 92

3. REPORT TYPE AND DATES COVERED
FINAL 1 Sep 88 to 31 Jul 92

4. TITLE AND SUBTITLE
COMPUTATIONAL AND NEURAL NETWORK MODELS FOR THE ANALYSIS
OF VISUAL TEXTURE

5. FUNDING NUMBERS
PE   61102F
PR   2313
TA   A8
AFOSR-88-0296

6. AUTHOR(S)
Drs R. Bajcsy and G. Gerstein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pennsylvania
School of Engineering and Applied Sciences
Dept of Computer and Information Science
200 South 23rd Street
Philadelphia PA 19104-6389

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR-

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research/NL
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
NOV 25 1992
S   E   D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The grant supported research efforts in two main research questions: (1) How do we
segment and estimate the slant/tilt of natural scenes with highly irregular textures, using
both biological and non-biological computing architectures; and (2) How do we design
machine perception and action systems and learning mechanisms that can improve their
performance through repeated feedback and interaction with their environment? Section
2 describes the research and publications relating to the slant/tilt and segmentation
problem. Section 3 describes research and publications using developmental
psychological and machine learning frameworks for learning visuomotor tasks. Finally,
Section 4 summarizes the two Ph.D dissertations in the Department of Computer and
Information Science made possible with AFOSR support.

14. SUBJECT TERMS

15. NUMBER OF PAGES
16

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
UNCLASSIFIED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)

14110

Final Report:
AFOSR Grant 88-0296

R. Bajcsy, G. Gerstein

October 1, 1992

The AFOSR grant was initiated on Sept. 1, 1988 and ended on July 31, 1992. This progress report describes activities that were partially supported by the grant during that period.

1 Overview

The grant supported research efforts in two main research questions:

1. How do we segment and estimate the slant/tilt of natural scenes with highly irregular textures, using both biological and non-biological computing architectures?

2. How do we design machine perception and action systems and learning mechanisms the can improve their performance through repeated feedback and interaction with their environment?

Section 2 describes the research and publications relating to the slant/tilt and segmentation problem. Section 3 describes research and publications using developmental psychological and machine learning frameworks for learning visuomotor tasks. Finally, section 4 summarizes the two Ph.D. dissertations in the Department of Computer and Information Science made possible with AFOSR support.

92

92-30193

13 OCT 1992

2 Research in Visual Texture Processing using Gabor Representations

The computational and neurobiological aspects of visual analysis using Gabor representations were investigated by the use of neural-network and high-fidelity neuronal simulations which explored the computational mechanisms for processing visual texture cues as well as the biological realizations of these computations.

The segmentation and slant/tilt problem was addressed through the utilization of the Gabor Filter as a signal measurement tool for images. Gabor functions were chosen because of their physiological plausibility and their excellent properties from a signal theoretic standpoint.

2.1 The Slant/Tilt Problem

The slant/tilt problem consists of estimating the inclination of a uniformly textured plane in three-dimensional space under the perspective imaging transformation. This transformation leads to distance and foreshortening effects which cause systematic distortions in spatial frequencies in the image plane. Gabor filters provide an ideal signal measurement tool for these cues because they minimize the positional and frequency joint uncertainty relation. These systematic changes provide the cues which allow the mapping to be partially inverted for determining the underlying inclination.

The engineering aspects of the slant/tilt estimation problem were addressed by the design of a parallel cooperative algorithm with excellent accuracy and robustness (see Page 13 for abstract). By using a relaxation algorithm and an ensemble of Gabor filters for local spectral measurements, natural scenes with a high-degree of irregularity and noise were able to be processed, significantly advancing the state of the art in texture gradient based scene analysis. Aside from the theoretical advances in this work, its intrinsic parallel formulation makes it a natural candidate for implementation with massively parallel computer hardware.

Insight into psychophysical aspects of texture perception were advanced by this cooperative algorithm. By analyzing the error properties of this algorithm, and the process of texture gradient measurement using Gabor filters, an explanation of psychophysical errors relating to the underestimation of

slant/tilt by human observers was devised (See abstract on Page 11). This result led to new insights regarding how the human visual system processes visual cues.

A subsystem which *learned* to estimate slant/tilt of an inclined plane from a set of training images was also developed. This system used an ensemble of Gabor filters to make local spectral measurements at different points in the image and then *learned* a mapping between these measurements and the inclination estimates using a back-propagation three layer neural network (This work is described on Page 13). By examining the middle layer units of the trained network, a diagonally oriented receptive field, reminiscent of the Hubel-Wiesel spatial receptive fields was found. However, the receptive field was expressed in the joint spatio-frequency space. This work suggested that diagonally selective receptive fields can be used for a variety of higher order visual computations, including texture gradient processing, rather than just feature detection in the spatial domain.

2.2 Texture-Based Segmentation

The earlier work of Turner [8] described and tested a parallel architecture for the segmentation of scenes using Gabor filter ensembles as measurement tools for aggregating image regions. A series of projects on texture discrimination using networks of laterally connected spiking elements (i.e. somewhat closer than usual to neural reality) have been completed.

This work was done using a highly realistic neuronal simulation to test the feasibility that a lateral interaction approach for segmentation as described in Turner's work [8] might be that which is used in cortical systems (This work is described on Page 3). The resulting network has been able to successfully segment images, and does so in approximately the same time course that occurs in human textural segmentation.

2.3 Parallel Algorithms for Gabor Ensemble Applications

All of the above algorithms require a large number of Gabor filters to be applied to images. We have developed a parallel implementation of the Gabor Filter Convolution operation on the Thinking Machines Corporation CM-1

computer with 4096 processors. The algorithm allows for $O(n)$ convolution of Gabor kernels with images, where n is the width of the Kernel Mask, rather than $O(nm^2)$ on a parallel machine. This allows the rapid application of ensembles of Filters, which was previously a bottleneck in the testing of the algorithms being developed.

3   Learning for Perception-Action

An important problem facing robotics is the design of systems that adapt and improve their performance based on their perceptual and motor interactions with their surroundings.

There are numerous issues relating to learning. The first major issue that relates to learning from examples is that of dimensionality. That is, the higher the dimensionality of the input and output space, the more learning examples necessary. This is known more formally in approximation theory as the "curse of dimensionality", which states that the number of exemplars necessary to approximate a function to a given level of accuracy increases exponentially with the dimensionality of the input space of the function. This puts an extraordinary burden on learning systems for large amounts of data in order to learn successfully. Recently, the statistical community has come forward with algorithms for non-parametric statistical regression which provide relief from this "curse" for certain classes of functions. In particular, the Projection Pursuit Approach [1] decomposes very high-dimensional functions into an additive expansion of projections of the higher dimensional space. This approach is equivalent to a two layer network with adjustable weights and unit functions. Recently, this projection pursuit expansion has been proven to avoid the curse of dimensionality for a broad class of functions [9]. We have successfully applied this Projection Pursuit Learning Approach to the problem of learning to grasp objects (See the abstracts on pages 10 and 8) in the plane using a robotic grasper and visual scene representation.

Additionally, a major problem with supervised approaches to learning occurs when a learning system is presented with a non-convex forward mapping in which a single input maps to several feasible outputs, all of which are present in the learning set. If the map is non-convex, the network may attempt to average the multiple outputs, which may yield a value outside of the feasible set. In this case, all solution alternatives must be stored ex-

4

plicitly or the learning network must censor all alternatives in the learning set except for one. We developed a variable resolution tree-based approach that allows for the representation of multiple alternative action maps. This approach can be used with a variety of methods.

A density adaptive learning (DARLING) algorithm was developed for classification tasks (see abstract on Page 7). The method allocates representational nodes in a density adaptive fashion that allows for efficient allocation of units based on the local density of exemplars in a given region of the parameter space using k-D-trees [2] combined with local classification trees. This algorithm has been applied to simulations of three-dimensional grasping tasks.

The DARLING algorithm shares many properties with neural-network algorithms such as Kohonen's self-organizing networks. The k-D tree is a probability equalizer; it takes an arbitrary probability distribution and maps it into a uniform distribution by a local partitioning of the parameter space. This metric scaling is interesting from several perspectives. If we think of this local scaling as selective magnification of the parameter space, then we have better discriminability where we have more exemplars. This type of property is often seen in human discrimination abilities where better discrimination is possible between similar objects that have been repeatedly encountered and have fine "nuance" differences between them when compared to those that are infrequently encountered and quite different [6]

This local metric scaling property has also been seen in the formation of somatotopic tactile maps in the monkey cortex [4] and in computer simulations of these structures [5]. Cortical structures that represent sensory information generally obey a topographic mapping property: areas close to each other on the sensing sheet are close to each other on the cortical sheet in which they are represented. Merzenich *et al.* demonstrated that the area of cortex that represents a given region of skin sensors increases in size in proportion to the amount of stimulation that the given region receives. This type of neuronal resource allocation allows for more neurons to be allocated where more data is available. The changes in cortex can take place in only a few hours, which is quite a rapid time scale. Kohonen [3] has also developed and investigated self-organizing networks with this same density adaptive property, and they have been utilized as associative memories for motor control of simulated robots [7].

In order for a learning system to be adaptive to changes in the world,

5

it must have some mechanism for forgetting observations. In neural network systems this is accomplished by weight changes that are accomplished through Hebbian learning mechanisms. The DARLING algorithm is a member of the class of memory-based learners, meaning that specific instances are held in memory and structured in a form that allows them to be used to predict future instances. However, this implies that in order to track the environment, obsolete memories, which no longer reflect the actuality of the world, must be deleted in an efficient and timely fashion so as not to bias the learner's prediction. A density adaptive forgetting algorithm which selectively deletes exemplars based on the amount of succeeding local evidence has been developed which allows memory-based learning systems to track and adapt to changes in the environment. This algorithm has been tested in simulation by adapting successfully both the motor and perceptual failures. This forgetting algorithm is general purpose and can be applied as a front end to a variety of learning systems, including neural networks which are often used in a batched mode.

4 Ph.D. Dissertations

The AFOSR grant 88-0296 has supported the training and dissertation research of two Ph.D. students, Marcos Salganicoff ("Learning and Forgetting for Perception-Action: A Projection Pursuit and Density Adaptive Approach" 9/92) and Mark Turner ("Local Spectral Models for Texture Segmentation and Estimation of Textured Surface Inclination" 7/90), in the Department of Computer and Information Science.

A Publication Titles and Abstracts

The section lists publications resulting directly from AFOSR support during the grant period.

1. "Learning and Forgetting for Perception-Action: A Projection Pursuit and Density Adaptive Approach," Marcos Salganicoff, Dissertation, Department of Computer and Information Science, September 1992.

We study learning of perception-action relations using visually-driven grasping as an example task.

The well-established technique of non-parametric Projection Pursuit Regression (PPR) is used to accomplish reinforcement learning by searching for projections of high-dimensional data sets that capture invariants in the distribution of reinforcement in the parameter-space. The variable resolution 2^k-tree, a generalized quadtree, is used to represent perception-action maps *based on* the resulting reinforcement regression function.

We also pursue the following problem: how can we use human expertise and insight into grasping to train a system to select gripper approach directions and orientations for grasping, and then have it verify and adapt its skills through trial and error? To accomplish this learning we develop a new *Density Adaptive* Reinforcement Learning algorithm. This algorithm uses statistical tests to identify regions of the attribute space in which the dynamics of the task change and the density of exemplars is high. It concentrates the building of high-resolution descriptions in those areas.

In order to adapt the default rules to those necessary for the robot, it is necessary for the system to be able to forget previous experiences that no longer reflect the behavior of the world. A general purpose *Density Adaptive* forgetting algorithm has been developed that can be used as a front-end for a variety of learning methods. Additionally, by setting the forgetting parameters appropriately, an upper bound on

the number of exemplars stored in the system may also be selected. This is important since all memory-based learning systems have finite memory in practice.

The approach is tested through simulation and experimentation. A robotic system incorporating two robots with a gripper, compliant instrumented wrist, arm, camera and laser scanner is used for experimentation. Since trial and error learning processes imply that failures will occur, the mechanics of the untrained robotic system must be able to tolerate mistakes during learning and not be damaged by excessive forces. We address this by the use of an instrumented, compliant robot wrist that controls impact forces.

2. "Robotic Sensorimotor Learning in Continuous Domains," Marcos Salganicoff and Ruzena Bajcsy, IEEE International Conference on Robotics and Automation, Nice, France, May, 1992.

We use findings in developmental psychology, neurophysiology, and machine perception to guide a robotic learning system's level of representation both for actions and for percepts. Visually driven grasping is chosen as the experimental task since it has general applicability and it has been extensively researched from several perspectives. An implementation of a robotic system with a dexterous three fingered hand, compliant instrumented wrist, arm and vision is used to test these ideas. Several sensorimotor primitives are implemented in this system and may be thought of as the "innate" perceptual and motor abilities of the system.

Applying empirical learning techniques to real situations brings up some important issues such as observation sparsity in high-dimensional spaces, arbitrary underlying functional forms of the reinforcement distribution and robustness to noise in exemplars. The well-established technique of non-parametric projection pursuit regression (PPR) is used to accomplish reinforcement learning by searching for projections of high-dimensional data sets that capture task invariants. Additionally, the learning process generally implies

8

failures along the way. Therefore, the mechanics of the untrained robotic system must be able to tolerate mistakes during learning and not be damaged. We address this by the use of an instrumented compliant robot wrist that controls impact forces.

3. "A Neural Network Model For Texture Discrimination," Jing Xing and George L. Gerstein, submitted to Biological Cybernetics, April 1992.

A model of texture discrimination in visual cortex was built using a feedforward network with lateral interactions among relatively realistic spiking neural elements. The elements have various membrane currents, equilibrium potentials and time constants, with action potentials and synapses. The model is realized by modified programs of MacGregor (1987). Gabor-like filters are applied to overlapping regions of the original image; the neural network with lateral excitatory and inhibitory interactions then compares and adjusts the Gabor amplitudes in order to produce the actual texture discrimination. Finally a combination layer selects and groups various representations in the output of the network to form the final transformed image material.

We show that both texture segmentation and texture boundary detection can be represented in the firing activity of such a network for a wide variety of synthetic to natural images. Performance details depend most strongly on the global balance of strengths of the excitatory and inhibitory lateral interconnections. The spatial distribution of lateral connective strengths has relatively little effect. Detailed temporal firing activities of single elements in the lateral connected network were examined under various stimulus conditions. Results show (as in area 17 cortex) that a single element's response to image features local to its receptive field can be altered by changes in the global context.

4. "Sensorimotor Learning Using Active Perception in Continuous Domains," AAAI Fall Symposium on Sensory Aspects of Robot Intelligence, Asilomar, Ca., November 14-17, 1991. pp. 78-85,

We propose that some aspects of task based learning in robotics can be approached using nativist and constructivist views on human sensorimotor development as a metaphor. We use findings in developmental psychology, neurophysiology, and machine perception to guide a robotic learning system's level of representation both for actions and for percepts. Visually driven grasping is chosen as the experimental task since it has general applicability and it has been extensively researched from several perspectives. An implementation of a robotic system with a dexterous three fingered hand, compliant instrumented wrist, arm and vision is used to test these ideas. Several sensorimotor primitives (vision segmentation and manipulatory reflexes) are implemented in this system and may be thought of as the "innate" perceptual and motor abilities of the system.

Applying empirical learning techniques to real situations brings up some important issues such as observation sparsity in high dimensional spaces, arbitrary underlying functional forms of the reinforcement distribution and robustness to noise in exemplars. The well established technique of non-parametric projection pursuit regression (PPR) is used to accomplish reinforcement learning by searching for *generalization directions* determining projections of high dimensional data sets which capture task invariants. Additionally, the learning process generally implies failures along the way. Therefore, the mechanics of the untrained robotic system must be able to tolerate grave mistakes during learning and not damage itself. We address this by the use of an instrumented compliant robot wrist which controls impact forces.

10

5. "HEAP: A Sensory Driven Distributed Manipulation System," Sanjay Agrawal, Marcos Salganicoff, Michael Chan, Ruzena Bajcsy, *Fifth International Conference on Advanced Robotics, Robots in Unstructured Environments*, pp. 649-654, June 19-22, 1991, Pisa, Italy, also published as MS-CIS-90-90, GRASP LAB 243.

> We address the problems of locating, grasping, and removing one or more unknown objects from a given area. In order to accomplish the task we use HEAP, a system of coordinating the motions of the hand and arm. HEAP also includes a laser range finer, mounted at the end of a PUMA 560, allowing the system to obtain multiple views of the workspace.
>
> We obtain volumetric information of the objects we locate by fitting superquadric surfaces on the raw range data. The volumetric information is used to ascertain the bust had configuration to enclose and constrain the object stably. The Penn Hand used to grasp the object, is fitted with 14 tactile sensors to determine the contact area and the normal components of the grasping forces. In addition the hand is used as a sensor to avoid any undesired collisions. The objective in grasping the objects is not to impart arbitrary forces on the object, but instead to be able to grasp a variety of objects using a simple grasping scheme assisted with a volumetric description and force and touch sensing.

6. "Underestimation of textured surface inclination by human observers: A model," *Biological Cybernetics*, 65(4), 1991, pp. 215-226.

> Since the experiments of Gibson in the 1950's it has been known that, with monocular stationary views, human subjects in psychophysical experiments will tend to underestimate the inclination of textured surfaces. This is particularly true when the textures are perceived as being "irregular". Two dimensional distributions of Gabor amplitudes from images of such surfaces are frequently altered in ways consistent with regular textures at lower angles of inclination. The cooperative algorithm described in the annual re-

port exhibits the same underestimation behavior as humans, suggesting that a similar computation may be occurring in visual cortex.

7. "Neural Assemblies as Building Blocks of Cortical Computation," George L. Gerstein and Mark R. Turner. In: Computational Neuroscience, Ed: Eric. L. Schwartz, MIT Press, Cambridge, Mass. 1990.

> This paper reviews experimental observations on neuronal assemblies from a number of laboratories and preparations. The principal finding is that correlation structure of firing (which can be interpreted as assembly organization) is highly dynamic, and depends on context. A second part of the paper describes a cooperative and iterative algorithm for texture segmentation. Several examples of the dynamic "condensation" of the model are given. The evolving weight structure between elements of the model is reminiscent of the dynamic weight structure in real neuronal assemblies, both with respect to the final state and its development.

8. "Estimation of textured surface inclination by parallel local spectral analysis," Mark Turner, Ruzena Bajcsy and George Gerstein, MS-CIS-89-42/Grasp Lab Report 184,

> This paper describes a cooperative algorithm for estimating the inclination of planar textured surfaces from the two dimensional distributions of Gabor filter amplitudes. An analytic model is first developed for these two dimensional distributions. Using this model, a hill climb algorithm is able to estimate the inclination of textured surfaces. The algorithm may be viewed as operating in parallel on a number of patches of the image and consolidating a global inclination value by lateral propagation of local estimates between regions.

9. "Local Spectral Analysis of Texture Gradients," Mark R. Turner. Presented at the Optical Society of America annual meeting, Orlando, Florida, October, 1989.

> The perspective image of an obliquely inclined textured surface exhibits shape and density distortions of texture elements. These distortions also systematically shift the projection of the spatial frequencies of which the texture is comprised. Using a set of filters with suitable spatial, frequency and orientation resolution, the inclination angle of many textured surfaces may be estimated from these frequency shifts. An algorithm has been developed which uses the amplitude distributions of 2D Gabor filters to perform this calculation on planar surfaces. The algorithm may be viewed as operating in parallel on a number of patches of the image and consolidating a global inclination value by lateral propagation of local inclination estimates between regions. This model provides a basis for the investigation of certain psychophysical aspects of the problem as well as possible physiological mechanisms (such as higher order receptive fields in simulated and biological neural networks) which underlie the perception of texture gradients.

10. "Learning to Compute Slant of Uniformly Textured Surfaces using Gabor Filters and Receptive Fields," Marcos Salganicoff. Presented at the Optical Society of America annual meeting, Orlando, Florida, October, 1989.

> A three layer non-recurrent network of simple non-linear elements can learn to compute the slant of a set of uniformly textured planar surfaces. During the learning phase, input training images are processed with Gabor filters as present in early mammalian vision and these filter responses are presented simultaneous with the known output slant values. Connection weights are adjusted to minimize error between the computed slant and the actual slant using an iterative gradient descent learning algorithm. After convergence, the

operational phase begins, in which the network is used with the weights fixed at values determined during learning to compute the slant of input images. The network is shown to achieve a excellent correlation with the actual slant of the surface.

The weights between the first and second layer units of the network can be interpreted biologically as receptive fields. The qualitative characteristics of some of the fields observed are oriented diagonal in nature, similar to those seen in Visual Area 17, but in a novel multi-dimensional space consisting of spatial frequency, orientation and image location. The implication to neurophysiological and machine vision is that oriented receptive fields in this space can serve well in the measurement of texture gradient.

References

[1] J. H. Friedman. *Classification and Multiple Regression Through Projection Pursuit.* Technical Report LCS 12, Stanford University, Laboratory for Computational Statistics, Stanford, CA, 1985.

[2] J. H. Friedman, J. L. Bentley, and R. A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3:209–226, 1977.

[3] T. Kohonen. *Self-Organization and associative memory.* Volume 8, Spring Series in Information Sciences, Heidelberg, 1984.

[4] M.M. Merzenich, R.J. Nelson, J.H. Kaas, M.P. Stryker, W.M Jenkins, J.M. Zook, M.S. Cynader, and J.M. Zook. Variability in hand surface representations in areas 3b and 1 of the adult owl and squirrel monkey. *Journal of Comparitive Neuroscience*, 25:281–296, 1987.

[5] J. C. Pearson, L. H. Finkel, and G. M. Edelman. Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection. *Journal of Neuroscience*, 7(12):4209–4223, 1987.

[6] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction.* Addison-Wesley, Reading, Massachussetts, 1991.

[7] H. Ritter, T. Martinetz, and K. Schulten. Topology conserving map for learning visuomotor coordination. *Neural Networks*, 2:159–168, 1989.

[8] M.R. Turner. Texture discrimination by gabor functions. *Biological Cybernetics*, 71–82, 1986.

[9] Y. Zhao and C. Atkeson. Some approximation properties of projection pursuit learning networks. In *Neural Information Processing Systems 4*, pages 936–943, Morgan-Kauffman, 1991.